BEST AVAILABLE COPY

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2002-157091
(43)Date of publication of application : 31.05.2002

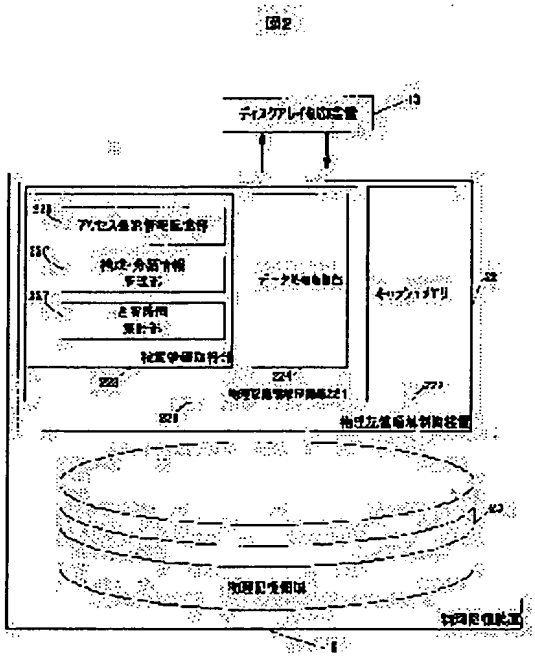
(51)Int.Cl. G06F 3/06
G06F 12/16

(21)Application number : 2000-353010
(22)Date of filing : 20.11.2000

(71)Applicant : HITACHI LTD
(72)Inventor : EGUCHI KENTETSU
MOGI KAZUHIKO
ARAKAWA TAKASHI
OEDA TAKASHI
ARAI HIROHARU

(54) STORAGE SUB-SYSTEM, AND MEMORY USED THEREFOR

(57)Abstract:
PROBLEM TO BE SOLVED: To obtain an occupied time of a logic storage area in a physical memory, and to obtain precise access occupied time information in every I/O to the physical memory.
SOLUTION: A physical storage area controller 22 on the individual physical memory 15 is provided with a table 225 for storing information about access requirement from a host computer, a table 227 for totalizing the occupied time as to access, a table 226 for control information for classifying constitution of a disk array, and a data processing control part 224 for obtaining constitution information and classification information of the logic storage area form a disk array controller 13, and for requesting the constitution information and the classification information of the logic storage area to the disk array controller, when necessary. The disk array controller 13 is provided with a means for transmitting the constitution information of the disk array at the present time to the physical storage area controller in response to the request from the physical storage area controller on the physical memory.



LEGAL STATUS

[Date of request for examination] 18.07.2003
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number]
[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] A means to be connected to 1 or two or more computers, and to acquire the operating condition information on two or more physical memory equipments and two or more of these physical memory equipments, In the storage subsystem which has a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment It is the storage subsystem which each of two or more of said physical memory equipments is equipped with a physical memory field control unit, and is characterized by this physical memory field control unit having a means to acquire the operating condition of a physical memory field.
[Claim 2] A means to be connected to 1 or two or more computers, and to acquire the operating condition information on two or more physical memory equipments and two or more of these physical memory equipments, In the storage subsystem which has a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment A means to acquire the operating condition information on said two or more physical memory equipments, and a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment It is prepared in the control unit which controls two or more above-mentioned physical memory equipments. Said control unit Furthermore, it has a means to transmit the information which performed matching with the logic storage region of physical memory equipment, and the physical memory field of physical memory equipment to said two or more physical memory equipments of each. It is the storage subsystem which each of two or more of said physical memory equipments is equipped with a physical memory field control unit, and is characterized by this physical memory field control unit having a means to acquire the operating condition of a physical memory field.
[Claim 3] It is physical memory equipment which is equipped with a physical memory field control unit in the physical memory equipment which constitutes a storage subsystem, and is used for the storage subsystem according to claim 1 or 2 characterized by this physical memory field control unit having a means to acquire the operating condition of a physical memory field.
[Claim 4] Said physical memory field control unit is physical memory equipment according to claim 3 characterized by having further a means to store the operating condition information on the acquired physical memory field.
[Claim 5] Said physical-memory field control unit is physical-memory equipment according to claim 3 or 4 characterized by to have further a means store the information which matched the logic storage region and the physical memory field of the physical-memory equipment which receives the operating condition information on the physical memory field of self-physical memory equipment from a means to transmit to said control unit, and said control unit, according to the acquisition demand of the operating-condition information on a physical-memory field which receives from said control unit.

[Translation done.]

* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[Field of the Invention] This invention relates to the store used for a storage subsystem and its system, and relates to the store used for the storage subsystem which has two or more stores especially, and its system.
[0002]
[Description of the Prior Art] The disk array system is known as a highly efficient secondary-storage system used for a computer system.
[0003] A disk array system is a system which made it possible to perform read/write of the data which arrange two or more physical memory equipments in the shape of an array, divide and store data in each physical memory equipment, and said each physical memory equipment is operated to juxtaposition, and are stored in said each physical memory equipment by dividing at a high speed.
[0004] As a conventional technique about a disk array system, the technique indicated by D.A.Patterson, G.Gibson, and R.H.Kats, "A Case for Redundant Arrays of Inexpensive Disks (disk array) etc." (inch Proc.ACM SIGMOD, pp.109-116, and June 1988), etc. is known. This conventional technique gives the classification of level 5 from level 1 to the disk array system which added redundancy according to that configuration. Moreover, a disk array system without redundancy may be added to these classification, and this may be called level 0 to them. In order to realize each above-mentioned level as a different configuration according to redundancy etc., cost differs from performance characteristics etc. And in building a disk array system, the array (group of physical memory equipment) of two or more level is made intermingled in many cases. Here, the group of the disk array which added redundancy is called a parity group. Moreover, in order to realize optimal cost performance in cost's changing with the engine performance, capacity, etc. also about physical memory equipment, and building a disk array system, two or more sorts of physical memory equipments with which the engine performance differs from capacity too may be used.
[0005] It distributes to said above physical memory equipments, and the data stored in a disk array system are arranged. For this reason, a disk array system needs to perform matching of the physical memory field which shows the logic storage region which the host computer connected to a disk array system accesses, and the storage region of said physical memory equipment, i.e., address translation.
[0006] The technique indicated by JP,9-274544,A etc. is known as a conventional technique about the disk array system which processes address translation. This conventional technique says that a means to acquire the information about I/O access over the logic storage region from a host computer, and a means to change matching with the physical memory field of a logic storage region, and to perform physical relocation realize the optimal arrangement of the stored data, and the technique in which a disk array control unit acquires the I/O access occupancy hour entry to a logic storage region is indicated by this official report.
[0007] In order to perform the load distribution of a disk array system, in case it changes matching with the physical memory field of a logic storage region and performs physical

relocation, since the I/O access occupancy hour entry to a logic storage region serves as data which become origin, it is important.
[0008]
[Problem(s) to be Solved by the Invention] Although, as for the conventional technique indicated by the official report mentioned above, it is said that a disk array control unit acquires the occupancy time amount by I/O to a logic storage region, the approach shown in this conventional technique has the trouble that it explains below.
[0009] First, the case where the writing (light) of data is performed to a certain logic storage region is considered. In this case, the light of data is performed to the physical memory field corresponding to the logic storage region concerned. A physical memory field is in physical memory equipment, and physical memory equipment is constituted by the cache memory and the physical memory field which mainly carry out the cache of a physical memory field control section and the data. And when carrying out the light of the data to a physical memory field, it writes in, when a physical memory field control section writes light data in a cache, and the response of termination is notified to a disk array control unit. For this reason, in order that the conventional technique mentioned above may actually carry out the light of the data, it will have the trouble that the time amount which accessed the physical memory field is not known.
[0010] Next, the case where reading (lead) of data is performed to a certain logic storage region is considered. In this case, in fact, although the lead of the data in the physical memory field corresponding to the logic storage region concerned is performed, when that lead data is in the cache memory in physical memory equipment, a physical memory field is not accessed, but cache memory is accessed, and that data is returned. For this reason, it will have the trouble that the exact time amount which could not distinguish and accessed the physical memory field does not understand whether the conventional technique mentioned above actually had access in the physical memory field for the data lead.
[0011] Moreover, the case where the logic storage region currently crossed to A, B, C, D, and multiple times has access is considered. And the response of A' and B is made into B' etc. for the response of A, and time of day of B (t) and response A' is made [the time of day of Access A] into A' (t) etc. for the time of day of A (t) and Access B. Here, Access A accesses physical memory equipment with a data lead, leads data, and assumes Access B to be what had the response B' without accessing physical memory equipment, since data are in a cache with a data lead. In this case, the direction of response B' of Access B which came after response A' of the access A which suited previously becomes previously. That is, it becomes A(t) <B (t), A' (t)> B' (t). At this time, the conventional technique which acquires the occupancy time amount by I/O to a logic storage region in a disk array control unit produces the trouble that sweet red bean soup with mochi cannot perform whether which I/O accessed the physical memory field in physical memory equipment how much, or it hit into the cache of physical memory equipment.
[0012] The purpose of this invention is by solving the trouble of the conventional technique mentioned above and acquiring the occupancy time amount of a logic storage region with physical memory equipment to offer the store which offers the storage subsystem which enabled it to acquire the occupancy time amount (real operating time) of each logic storage region by the system configuration unacquirable [with a disk array control unit], and is used for this.
[0013] The purpose of this invention moreover, by taking into consideration the effect on the physical memory equipment for every I/O In order to be able to make small the error of the analysis of the utilization factor of the logic storage region of said disk array system, or utilization factor prediction and to perform optimum performance tuning more It is in offering the store which offers the storage subsystem which can acquire the access occupancy hour entry for every I/O to physical memory equipment with a more high precision, and is used for this.
[0014]
[Means for Solving the Problem] According to this invention, said purpose is connected to 1 or two or more computers. Two or more physical memory equipments, In the storage subsystem which has a means to acquire the operating condition information on two or more of these physical memory equipments, and a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical

memory equipment A means to acquire the operating condition information on said two or more physical memory equipments, and a means to perform matching with the logic storage region which said computer makes a read/write object, and the physical memory field of said physical memory equipment It is prepared in the control unit which controls two or more above-mentioned physical memory equipments. Said control unit Furthermore, it has a means to transmit the information which performed matching with the logic storage region of physical memory equipment, and the physical memory field of physical memory equipment to said two or more physical memory equipments of each. It has a physical memory field control unit, and this physical memory field control unit is attained by having a means of two or more of said physical memory equipments to acquire the operating condition of a physical memory field, respectively.
[0015] Moreover, said purpose is set to the physical memory equipment which constitutes a storage subsystem. A means by which have a physical memory field control unit and this physical memory field control unit acquires the operating condition of a physical memory field, An acquisition demand of the operating condition information on a means to store the operating condition information on the acquired physical memory field, and the physical memory field received from said control unit is accepted. It is attained by having a means to transmit the operating condition information on the physical memory field of self-physical memory equipment to said control unit, and a means to store the information which matched the logic storage region and physical memory field of the physical memory equipment received from said control unit.
[0016]
[Embodiment of the Invention] Hereafter, a drawing explains the operation gestalt of the storage used for the storage subsystem by this invention, and its system to a detail.
[0017] The block diagram showing the configuration of the computer system with which drawing 1 was equipped with the storage subsystem by this invention, and drawing 2 are the block diagrams showing the configuration of physical memory equipment. In drawing 1 R> 1 and drawing 2 a host and 12 10 A storage subsystem, 13 disk array control information and 15 for a disk array control device and 14 Physical memory equipment, In 16, a disk array and 17 an I/O bus and 19 for a control terminal and 18 A network, 22 a physical memory field and 130 for a physical memory field control unit and 23 The read/write processing section, 131 the relocation decision processing section and 133 for the operating condition information acquisition processing section and 132 The relocation executive operation section, 141 class configuration information and 143 for the information corresponding to logic/physics, and 142 Class attribute information, 144 physical field operating condition information and 146 for logic field operating condition information and 145 Relocation decision horizon information, 147 free-space information and 149 for relocation activation time information and 148 Relocation information, 14A -- a storage occupancy hour entry and 221 -- for the operation information acquisition section and 224, as for the access request information storage section and 226, a data-processing control section and 225 are [a physical memory field control section and 222 / cache memory and 223 / configuration / classification Research and Data Processing Department and 227] the occupancy time amount total sections.
[0018] The computer system shown in drawing 1 consists of 1 which is the calculating machine of a high order or two or more hosts 10, a storage subsystem 12, and a control terminal 17. It connects with the storage subsystem 12 by I/O bus 18, and a host 10 publishes lead of data, and I/O for light processing to the storage subsystem 12. In case this I/O is performed, a host 10 specifies the logical storage region of the storage subsystem 12. That is, a host 10 accesses with the address of a usually logical storage region to the data in a storage subsystem. Moreover, I/O bus 18 is constituted by ESCON, SCSI, the fiber channel, etc.
[0019] The storage subsystem 12 consists of a disk array control unit 13 and two or more physical memory equipments 15. The disk array control unit 13 is equipped with the read/write processing section 130, the operating condition information acquisition processing section 131, the relocation decision processing section 132, and the relocation executive operation section 133, and these processing sections process read/write processing, operating condition information acquisition processing, relocation decision processing, relocation executive operation, etc. Moreover, the disk array control non-dense of the storage subsystem 12 holds the

information 141 corresponding to a logic storage region / physical memory field, the class configuration information 142, the disk array configuration information 1400 of class attribute information 143 grade, storage occupancy hour entry 14A of the logic field operating condition information 144 and physical field operating condition information 145 grade, the relocation decision horizon information 146, the relocation activation time information 147, the free-space information 148, and relocation information 149 grade. In addition, others and parity group information, RAID level information, etc. which were mentioned above may be included in the disk array configuration information 14 mentioned above. [information]
[0020] Moreover, the host 10, the disk array control unit 13, and the control terminal 17 are mutually connected by the network 19. A network 19 may be constituted by Ethernet (trademark), FDDI, the fiber channel, etc. A control terminal 17 is usually used in order to perform maintenance, management, etc. of the storage subsystem 12.
[0021] Moreover, in explanation of the operation gestalt of this invention, although it exists, respectively, since it is not important, the component which surely exists in computers, such as memory for performing processing which appears in a host 10, the disk array control unit 13, and a control terminal 17, respectively, and CPU, is not specified here.
[0022] The class division of two or more physical memory equipments 15 formed in the above-mentioned storage subsystem 12 is carried out for every engine performance of physical memory equipment, and they constitute the disk array 16 for every class. Moreover, although not shown clearly, two or more physical memory equipments are used, and the parity group is constituted here. And each of physical memory equipment 15 is constituted by the physical memory field control device 22 which controls the physical memory field 23 and this physical memory field 23 to be shown in drawing 2 , and various data are stored in the physical memory field 23.
[0023] Moreover, if the address of the physical memory field 23 seems to have mentioned above directly from a host 10, it does not break, but a host 10 accesses the data on two or more logical storage regions on two or more physical memory fields 23. Namely, a host 10 accesses by specifying a logic storage region to the data in the storage region of each physical memory equipment 15 in the storage subsystem 12.
[0024] The disk array control unit 13 is connected with two or more physical memory equipments 15. The lead and light processing instruction I/O which controlled two or more physical memory equipments 15, or were emitted by said host 10 Make the address of the logic storage region where the appointed data exist, and the address of a physical memory field with the address of the logic storage region correspond, transmit data I/O to suitable physical memory equipment 15, and if it is light processing The data transmitted by the host 10 are transmitted to physical memory equipment 15, and if it is lead processing, the data transmitted from physical memory equipment 15 are received, and it is processing transmitting to a host 10 etc.
[0025] The physical memory field control unit 22 which it has in physical memory equipment 15 is constituted by the physical memory field control section 221 and cache memory 222. Cache memory 222 has the quick rate of processing of the read/write of data compared with the physical memory field 23. And cache memory 222 is used as follows about the data about the lead or light instruction transmitted from the disk array control device 13. That is, in case the light data transmitted from the disk array control device 13 are written in the physical memory field 23 in light processing, data are written also in cache memory 222. Moreover, in case data reading appearance is carried out from the physical memory field 23 in lead processing, when it was written in cache memory 222, or the same data are in cache memory and it comes to physical memory equipment from the disk array control device 13 as a lead instruction to the data by former lead processing, the read data do not read the data from the physical memory field 23, but read it from cache memory 222. Thereby, the processing engine performance of physical memory equipment 15 can be improved.
[0026] The physical memory field control section 221 is mainly equipped with the operation information acquisition section 223 and the data-processing control section 224, and is constituted. The data-processing control section 224 receives the lead or light instruction of data transmitted from the disk array control device 13. And the data-processing control section

224 will read the lead data from cache memory 222, if cache memory 222 is accessed and the lead data exists in cache memory 222, when the received instruction is a lead instruction, if the data is not in cache memory 222, it will access the physical memory field 23, will read the lead data, and will transmit data to the disk array control device 13. Moreover, the data-processing control section 224 is after that, and writes the data in the physical memory field 23 at the same time it writes the data transmitted from the disk array control device 13 in cache memory 222, when the received instruction is a light processing instruction. You may not write light data in cache memory 222, but may also write them in the direct physical memory field 23.

[0027] The operation information acquisition section 223 is constituted by the access request information storage section 225, configuration / classification Research and Data Processing Department 226, and occupancy time amount total section 227 grade. When a logic storage region with the data specified by the I/O process, the physical memory field 23 where the data exists, or cache memory 222 is accessed, the above-mentioned data-processing control section 224 classifies the hour entry of the access for every (random access, sequential access, etc.) processing classification of an I/O process, and records it on the occupancy time amount total section 227. Moreover, from the disk array control device 13, the data-processing control section 224 receives information, such as correspondence information on the address of the logic storage region in a disk array, and the address of the physical memory field 23, and engine performance of physical memory equipment 15, and records it on configuration / classification Research and Data Processing Department 226. Furthermore, the data-processing control section 224 receives a lead of data or light instruction data transmitted from disk array control-device 13 grade, and records the instruction data on the access request information storage section 225 so that it may be possible to receive two or more I/O processes.

[0028] The logic storage region according to an I/O process within physical memory equipment 15 when physical memory equipment 15 has a configuration which was mentioned above, the physical memory field 23, Or it becomes possible to classify the hour entry at the time of access when accessing cache memory 222 for every (random access, sequential access, etc.) processing classification of an I/O process, and to record on the occupancy time amount total section 227. It classifies into a physical memory field to what whether time amount access was carried out and physical memory equipment 15 of cache memory 22 it hit according to an I/O process, and it becomes possible to total the occupancy time amount.

[0029] Drawing 3 is a flow chart explaining processing actuation of a disk array control device when a storage subsystem is started, and explains this hereafter.

[0030] (1) The disk array control unit 13 transmits the information 141 corresponding to the logic/physics which is the matching information on the address of the logic storage region in the physical memory field 23 in physical memory equipment 15, and the address of the physical memory field where the logic storage region actually exists, the class configuration information 142, and the disk array configuration information 14 of class attribute information 143 grade to the physical memory equipment 15 connected with self-equipment 13 at the time of starting of the storage subsystem 12 (steps 300 and 310).

[0031] (2) Next, the disk array control unit 13 receives the notice the physical memory equipment 15 sent from the physical memory field control unit 22 changed [the notice] to the accessible ready state, when physical memory equipment 15 becomes accessible by transmission of the information mentioned above. Things come and physical memory equipment 15 is in the condition of the initialization termination by the disk array configuration information 14 (step 320).

[0032] (3) Then, the disk array control device 13 receives data with more various what has transmitted host I/O of a lead or light processing to the storage subsystem 12 to the logic storage region in the storage subsystem 12, things which deliver an instruction and data with disk array control devices than a host 10 by I/O-bus 18 course (step 330).

[0033] (4) When host I/O is received as said received data, the disk array control unit 13 receives the lead or light demand to the logic storage region specified by host I/O, and asks for the logic storage region address and the corresponding address of the physical memory field 23 using the information 141 corresponding to the logic/physics which changes the address (logical

address) of the logic storage region into the address (physical address) of a physical memory field (steps 340 and 350).

[0034] (5) The disk array control device 13 specifies the address of the physical-memory field where predetermined data exist, and, in lead processing, transmits light data to the physical-memory equipment which transmits lead data for the physical memory equipment which has the above-mentioned physical address to lead data to read-out and a host 10, receives the light data transmitted by the host 10 in light processing, and has the physical address (360).

[0035] Drawing 4 is a flow chart explaining processing actuation of a disk array control device when correspondence of the address of a logic storage region and the address of a physical memory field changes, and explains this hereafter.

[0036] (1) By the change in physical memory equipment 15, change of RAID level, and a logic storage region moving the disk array control device 13 to the address of a physical memory field different from a certain physical memory **** address now etc. When it supervises that correspondence of the address of a logic storage region and the address of a physical memory field changed and is changeful, physical memory equipment 15 is received again. The information 141 corresponding to the logic/physics which is the matching information on the address of the logic storage region in the physical memory field 23 in physical memory equipment 15, and the address of the physical memory field where the logic storage region actually exists, The class configuration information 142 and the disk array configuration information 14 of class attribute information 143 grade are transmitted (step 3101).

[0037] (2) Next, the disk array control unit 13 receives the notice the physical memory equipment 15 sent from the physical memory field control unit 22 changed [the notice] to the accessible ready state, when physical memory equipment 15 becomes accessible by transmission of the information mentioned above. Things come and physical memory equipment 15 is in the condition of the updating termination by the disk array configuration information 14 (step 3201).

[0038] (3) Subsequent processing is performed like the case of steps 3300, 3400, 3500, and 3600 explained by drawing 3 (steps 3301, 3401, 3501, and 3601).

[0039] In addition, in steps 320 and 3201 mentioned above, as for the disk array control unit 13, ** does not need to receive the information that the physical memory field control unit 22 to physical memory equipment 15 changed in the accessible condition. In this case, what is necessary is just to perform access processing for performing predetermined processing to physical memory equipment 15, when it assumes that it is in accessible conditions, such as a lead and a light, to physical memory equipment 15 and the access directions to a certain storage region come for the disk array control unit 13 after a certain fixed time amount. Moreover, when there is no response into waiting and fixed time amount until it performs access processing in order to perform predetermined processing again when there is no response to predetermined access processing, or a response comes back, it is good also as a method which tells that to the module which issued the access directions to a certain storage region.

[0040] Moreover, the information 141 corresponding to logic/physics in the above-mentioned is information to which a logic storage region and a physical memory field are made to correspond. And the logical address is the address which shows the logic storage region which a host 10 uses in said read/write processing section 130. Moreover, a physical address is the address which shows the field on the physical memory equipment 15 with which data are actually stored, and consists of the physical memory device number and the address in physical memory equipment. A storage number shows each physical memory equipment 15. The address in storage is the address which shows the storage region within physical memory equipment 15.

[0041] Drawing 5 is a flow chart explaining processing actuation of the disk array control device 13 at the time of the disk array control device 13 reading the information in the operation information acquisition section 223 of physical memory equipment 15, and explains this hereafter.

[0042] (1) After the storage subsystem 12 is started, the disk array control device 13 initializes storage occupancy hour entry 14A, and transmits an acquisition demand of the access occupancy hour entry of the physical memory equipment 15 to two or more physical memory

equipments 15 connected after that (steps 371 and 372).
[0043] (2) Next, the disk array control unit 13 stores the access occupancy hour entry of reception and each physical memory equipment in storage occupancy hour entry 14A for an access occupancy hour entry from each physical memory equipment 15 (steps 373 and 374).
[0044] In addition, the timing of acquisition of the access occupancy information on the disk array control unit 13 mentioned above The method which reads the access occupancy information by access to physical memory equipment 23 from the module of others by host I/O, backup, etc. to a fixed time interval from the occupancy time amount total section 227 in each physical memory equipment 15, When an access occupancy hour entry acquisition demand is transmitted to the disk array control unit 13 from other modules (for example, a host 10 and a control terminal 17), it is various and dependent on a design.
[0045] The access occupancy hour entry acquired by the above-mentioned is recorded on the occupancy time amount total table in the disk array control device 13.
[0046] Drawing 6 is a flow chart explaining processing actuation of the physical memory field control device 22 in physical memory equipment 15, next explains this.
[0047] (1) The physical memory field control device 22 receives the disk array configuration information 14 which is information on the logical address and the physical address in the physical memory field 23 of the data 15 transmitted from the disk array control device 13 at the time of starting of the storage subsystem 12, i.e., physical memory equipment, such as correspondence, (steps 400 and 401).
[0048] (2) The physical memory field control unit 22 which received the disk array configuration information 14 performs creation initialization of configuration / classification information management table of configuration / classification Research and Data Processing Department 226 in the operation information acquisition section 223, or the occupancy time amount total table of the occupancy time amount total section 227 based on the information (step 402).
[0049] (3) After initialization processing of the table in step 402 is completed, in order to make this physical memory equipment 15 recognize an accessible thing, notify that initialization processing was completed to the disk array control unit 13. In addition, it is not necessary to transmit the information to which physical disk equipment changed in the accessible condition to the disk array control device 13. When the access directions to a certain storage region come to the physical memory field control unit 22 of physical memory equipment 15 after the fixed time amount, in this case, the physical memory field control unit 22 If it is in a condition accessible to the physical memory field 23 in order to perform predetermined processing Access the physical memory field 23, perform predetermined processing, and if impossible [whether processing predetermined in the condition of having stored access information in the access request information storage section 225, and having become accessible to the physical memory field is performed, and] or you may make it not receive a certain access directions to a storage region until it comes to resemble the physical memory field 23 in the accessible condition, in order to perform predetermined processing (step 403)
[0050] (4) After that, the physical memory field control device 22 waits to transmit host I/O, and a physical memory equipment operation information acquisition demand instruction or new disk array configuration information from the disk array control device 13, and receives it (step 404).
[0051] (5) At step 404, if host I/O is received from the disk array control unit 13, when the I/O judges lead processing or light processing and it is lead processing, the data-processing control section 224 will confirm whether the data which should be read exist in cache memory 22 (steps 405 and 406).
[0052] (6) When the data is read from cache memory 22 when the data exists in cache memory 22, and the data does not exist in cache memory 22 with the check of step 406, read the data from the physical memory field 23, and transmit data to the disk array control unit 13 (steps 407, 709, and 408).
[0053] (7) At step 405, when judged with host I/O being light processing, the data-processing control section 224 receives the light data transmitted by the host 10, and writes the light data in cache memory 22 (steps 410 and 411).
[0054] (8) And the data-processing control section 224 stores the above-mentioned light data in

the physical memory field 23 while notifying the notice of data write-in termination to the disk array control device 13 (steps 412 and 413).
[0055] The data-processing control section 224 to cache memory 22 after processing of step 408, or processing of step 413 (9) In access ** Or the JOB classification information on the information on whether the physical memory field 23 was accessed, a random lead, a sequential lead, etc., The access classification of the JOB classification information on the random read/write at the time of writing light data in the physical memory field 23, sequential read/write, etc. is recognized. The occupancy hour entry which accessed cache memory 22 or the physical memory field 23 is stored in the occupancy time amount total section 227 in the operation information acquisition section 223 for every access classification (steps 414 and 415).
[0056] (10) If new disk array configuration information is received from the disk array control device 13 at step 404, the data-processing control section 224 will rewrite the information in configuration / classification Research and Data Processing Department 226, the operation information acquisition section 223, corresponding to new disk array configuration information (steps 418 and 419).
[0057] (11) When a physical memory equipment operation information acquisition demand instruction is received from the disk array control device 13 at step 404, the data-processing control section 224 reads the access occupancy hour entry of the physical memory equipment 15 stored in the occupancy time amount total section 227 in the operation information acquisition section 223, and transmits it to the disk array control device 13 (steps 416 and 417).
[0058] In addition, it may be made to perform transmission to the disk array control device 13 from the physical memory equipment 15 of the occupancy hour entry in processing of step 417 mentioned above to the disk array control device 13 with a fixed time interval automatically from physical memory equipment 15. In this case, said physical memory equipment operation information acquisition demand instruction is not transmitted to physical memory equipment 13 from the disk array control unit 13.
[0059] Drawing 7 is drawing explaining the example of a configuration of the information 141 corresponding to logic/physics in the table for managing correspondence with the address of a logic storage region and the address of a physical memory field which are held in the disk array control unit 13.
[0060] The disk array control unit 13 has managed correspondence with the address of the logic storage region in the physical memory field 23 in two or more physical memory equipments 15 connected, and the address of the physical memory field in the logic storage region. Each of the logic storage region number 500 given to a specific logic storage region, the logical address 510, the physical address 520 by the storage number 521 with a physical storage region with the logic storage region and the address 522 of a physical storage region, the RAID level 530 that shows the engine performance of the physical-memory equipment 15, and the parity group number 540 to which the physical-memory equipment 15 belongs is matched, and the information 141 corresponding to the logic/physics used for this is constituted, as shown in drawing 7 . When processing of a lead, a light, etc. specifies the address of a logic storage region from a host 10, a control terminal 17, and other modules (for example, other disk array control units etc.) and it is accessed to the self-disk array control unit 13 by having such information 141 corresponding to logic/physics, the disk array control unit 13 can change the address of a logic field into the address of a physical storage region, and can perform read/write processing of data correctly to physical memory equipment 15.
[0061] Drawing 8 is drawing showing the example of the logic field operating condition information 144 stored in the disk array control unit 13, and the storage occupancy hour entry 141 of physical field operating condition information 145 grade. Such information is constituted as an occupancy time amount total table.
[0062] The disk array control device 13 reads periodically the access occupancy information on the physical memory field 23 of the physical memory equipment 15 by access from the module of others by host I/O, backup, etc. from the occupancy time amount total section 227 in each physical memory equipment 15, and records the access occupancy hour entry on the occupancy time amount total table in the receiving disk array control device 13. The example shown in

drawing 8 is every logic storage region number 601 and I/O. Occupancy time amount is totaled every JOB classification 602. I/O As a JOB classification, although 670 and a total of 680 are shown by the example of illustration at the time of the sequential lead 610, the sequential light data 620, the sequential light parity 630, the random lead 640, the random light parity 660, and a cache hit, it is I/O of further others. There may be JOB classification.

[0063] The occupancy time amount read from physical memory equipment 15 may be record by the accumulation value of not only the above-mentioned but the access occupancy time amount for every I/O, and a universal time amount value and a time amount value peculiar to a machine. Moreover, in the disk array control unit 13, the access occupancy time amount of each logic storage region or a physical memory field may be edited, and an occupancy hour entry table may newly be created based on the value which found the access occupancy time amount for every physical memory equipment and every parity group.

[0064] When a host 10 and control terminal 17 grade give the operation information acquisition demand of physical memory equipment to the disk array subsystem 12 by the above-mentioned, even if a host 10 and control terminal 17 grade access direct physical memory equipment 15 and do not acquire the access occupancy hour entry of a logic storage region or a physical memory field, they become possible [acquiring the access occupancy hour entry of the logic storage region and physical memory field from the disk array control unit 13].

[0065] Drawing 9 is drawing showing the example of a configuration of the table which manages matching with the address of a logic storage region and the address of a physical memory field which are stored in configuration / classification Research and Data Processing Department 226 in the operation information acquisition section 223 in physical memory equipment 15.

[0066] From the disk array control device 13, when the relation of the address of a logic storage region and the address of a physical-memory field which are produced by the time of starting of the storage subsystem 12, the change in physical-memory equipment, change of RAID level, migration of a logic storage region, etc. changes, the data-processing control section 224 in physical-memory equipment 15 receives the matching information on the logic storage region and physical-memory field, and stores it in a correspondence table as shows it to configuration / classification Research and Data Processing Department 226 at drawing 9 . This correspondence table is constituted by physical ADORE 7520 by the logic storage region number 700 given to a specific logic storage region, the logical address 710, and the storage number 721 with a physical storage region with that logic storage region and the address 722 of a physical storage region. Thereby, it can recognize which address logic storage region physical memory equipment 15 has where of the physical memory regional address in self-physical memory equipment 15.

[0067] Drawing 10 is drawing showing the example of a configuration of the table of the occupancy hour entry by access to the storage region by I/O accumulated and stored at the occupancy time amount total section 227 of the operation information acquisition section 223 in physical memory equipment 15.

[0068] This table is every logic storage region number 801 and I/O. Occupancy time amount is totaled every JOB classification 802. I/O As a JOB classification, in the example of illustration, although 870 and a total of 880 are shown at the time of the sequential lead 810, the sequential light data 820, the sequential light parity 830, the random lead 840, the random light parity 860, and a cache hit, there may be I/OJOB classification of further others.

[0069] If the data-processing control section 224 in physical memory equipment 15 has access in a store by I/O from a host etc., it will accumulate the occupancy time amount 890 in the occupancy time amount total section 227 every JOB classification 802 of access about each logic storage region 801 with access. This becomes possible to obtain the relation between the number of the logic storage region in physical memory equipment 15, the occupancy time amount by the access classification to the logic storage region, and the sum total occupancy time amount within a certain time amount within physical memory equipment 15.

[0070] According to the operation gestalt of this invention mentioned above, the occupancy hour entry of the storage for every I/O can be acquired, and acquisition of the occupancy hour entry of the storage by access to a storage region can be realized within physical memory equipment.

[0071] According to the operation gestalt of this invention, the thing of two or more physical

memory equipments which constitute a storage subsystem for which the occupancy time amount of the logic storage region of physical memory equipment is acquired becomes respectively possible by the above-mentioned, and the thing of each logic storage region to do for occupancy time amount (real operating time) acquisition becomes possible by the system configuration unacquirable [with a disk array control unit]. Moreover, in order according to the operation gestalt of this invention mentioned above to be able to make small the error of the analysis of the utilization factor of the logic storage region of said storage subsystem, or utilization factor prediction and to perform optimum performance tuning more by taking into consideration the effect on the physical memory equipment for every I/O, the access occupancy hour entry for every I/O to physical memory equipment with a more high precision is acquirable.

[0072]

[Effect of the Invention] As explained above, according to this invention, physical memory equipment can acquire the occupancy time amount of a logic storage region, and the occupancy time amount (real operating time) of each logic storage region can be acquired by the system configuration unacquirable [with a disk array control unit].

[0073] Moreover, according to this invention, since the access occupancy hour entry for every I/O to physical memory equipment is acquirable, it becomes possible to perform the analysis of the utilization factor of the logic storage region of a storage subsystem and the prediction of a utilization factor in consideration of the effect on the physical memory equipment for every I/O with a small error, and the engine performance of the more nearly optimal storage subsystem can be tuned up.

[Translation done.]

* NOTICES *

JPO and NCIPI are not responsible for any
damages caused by the use of this translation.

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is the block diagram showing the configuration of the computer system equipped with the storage subsystem by this invention.
[Drawing 2] It is the block diagram showing the configuration of physical memory equipment.
[Drawing 3] It is a flow chart explaining processing actuation of a disk array control device when a storage subsystem is started.
[Drawing 4] It is a flow chart explaining processing actuation of a disk array control device when correspondence of the address of a logic storage region and the address of a physical memory field changes.
[Drawing 5] It is a flow chart explaining processing actuation of the disk array control device at the time of a disk array control device reading the information on operation information acquisition circles of physical memory equipment.
[Drawing 6] It is a flow chart explaining processing actuation of the physical memory field control device in physical memory equipment.
[Drawing 7] It is drawing explaining the example of a configuration of the information corresponding to the logic/physics which manages correspondence of the address of the logic storage region and physical memory field which are held in the disk array control unit.
[Drawing 8] It is drawing showing the example of storage occupancy hour entries, such as logic field operating condition information stored in a disk array control unit, and physical field operating condition information.
[Drawing 9] It is drawing showing the example of a configuration of the table which manages matching of the address of the logic storage region and physical memory field which are stored in configuration / classification Research and Data Processing Department in physical memory equipment.
[Drawing 10] It is drawing showing the example of a configuration of the table of the occupancy hour entry by access to the storage region accumulated and stored at the occupancy time amount total section of the operation information acquisition section in physical memory equipment.
[Description of Notations]
10 Host
12 Storage Subsystem
13 Disk Array Control Unit
14 Disk Array Control Information
15 Physical Memory Equipment
16 Disk Array
17 Control Terminal
18 I/O Bus
19 Network
22 Physical Memory Field Control Unit
23 Physical Memory Field
130 Read/write Processing Section

131 Operating Condition Information Acquisition Processing Section
132 Relocation Decision Processing Section
133 Relocation Executive Operation Section
141 Information corresponding to Logic/Physics
142 Class Configuration Information
143 Class Attribute Information
144 Logic Field Operating Condition Information
145 Physical Field Operating Condition Information
146 Relocation Decision Horizon Information
147 Relocation Activation Time Information
148 Free-Space Information
149 Relocation Information
14A Storage occupancy hour entry
221 Physical Memory Field Control Section
222 Cache Memory
223 Operation Information Acquisition Section
224 Data-Processing Control Section
225 Access Request Information Storage Section
226 Configuration / Classification Research and Data Processing Department
227 Occupancy Time Amount Total Section

[Translation done.]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2002－157091
（Ｐ2002－157091Ａ）
(43)公開日　平成14年５月31日(2002.5.31)

(51)Int.Cl.'	識別記号	ＦＩ		テーマコード'(参考)
Ｇ06Ｆ　3/06	302	Ｇ06Ｆ　3/06	302Ａ	5Ｂ018
	301		301Ｘ	5Ｂ065
	304		304Ｎ	
	540		540	
12/16	320	12/16	320Ｌ	

審査請求　未請求　請求項の数５　ＯＬ　（全 16 頁）

(21)出願番号　特願2000－353010(Ｐ2000－353010)

(22)出願日　平成12年11月20日(2000.11.20)

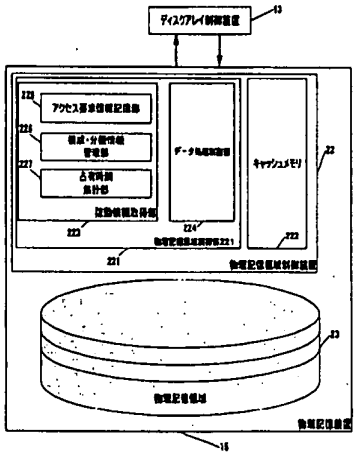
(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　江口　賢哲
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　茂木　和彦
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(74)代理人　100093492
弁理士　鈴木　市郎　（外１名）

最終頁に続く

(54)【発明の名称】　ストレージサブシステム及びそのシステムに使用する記憶装置

(57)【要約】
【課題】　物理記憶装置において、論理記憶領域の占有時間を取得し、より精度の高い物理記憶装置への１／０毎のアクセス占有時間情報を取得する。
【解決手段】　個々の物理記憶装置１５上にある物理記憶領域制御装置２２に、ホストからのアクセス要求に関する情報を記憶するテーブル２２５と、アクセスに関する占有時間を集計するテーブル２２７と、ディスクアレイの構成を分類する管理情報のテーブル２２６と、ディスクアレイ制御装置１３から論理記憶領域の構成情報や分類情報を取得し、必要に応じて論理記憶領域の構成情報や分類情報をディスクアレイ制御装置にリクエストするデータ処理制御部２２４を備える。また、ディスクアレイ制御装置１３に、物理記憶装置上にある物理記憶領域制御装置からのリクエストに応じて、現在のディスクアレイの構成情報を物理記憶領域制御装置に送信する手段を設ける。

図2



1

【特許請求の範囲】
【請求項１】　１または複数の計算機に接続され、複数の物理記憶装置と、これらの複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とを有するストレージサブシステムにおいて、前記複数の物理記憶装置のそれぞれは、物理記憶領域制御装置を備え、該物理記憶領域制御装置は、物理記憶領域の使用状況を取得する手段を有することを特徴とするストレージサブシステム。
【請求項２】　１または複数の計算機に接続され、複数の物理記憶装置と、これらの複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とを有するストレージサブシステムにおいて、前記複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とが、前述複数の物理記憶装置を制御する制御装置内に設けられ、前記制御装置は、さらに、物理記憶装置の論理記憶領域と物理記憶装置の物理記憶領域との対応付けを行った情報を前記複数のそれぞれの物理記憶装置に送信する手段を備え、前記複数の物理記憶装置のそれぞれは、物理記憶領域制御装置を備え、該物理記憶領域制御装置は、物理記憶領域の使用状況を取得する手段を有することを特徴とするストレージサブシステム。
【請求項３】　ストレージサブシステムを構成する物理記憶領域制御装置において、物理記憶領域制御装置を備え、該物理記憶領域制御装置は、物理記憶領域の使用状況を取得する手段を有することを特徴とする請求項１または２記載のストレージサブシステムに使用する物理記憶装置。
【請求項４】　前記物理記憶領域制御装置は、取得した物理記憶領域の使用状況情報を格納する手段をさらに備えることを特徴とする請求項３記載の物理記憶装置。
【請求項５】　前記物理記憶領域制御装置は、前記制御装置より受信する物理記憶領域の使用状況情報の取得要求に応じて、自物理記憶装置の物理記憶領域の使用状況情報を前記制御装置に送信する手段と、前記制御装置より受信する物理記憶装置の論理記憶領域と物理記憶領域とを対応付けした情報を格納する手段をさらに備えることを特徴とする請求項３または４記載の物理記憶装置。
【発明の詳細な説明】
【０００１】
【発明の属する技術分野】本発明は、ストレージサブシステム及びそのシステムに使用する記憶装置に係り、特に、複数の記憶装置を有するストレージサブシステム及びそのシステムに使用する記憶装置に関する。
【０００２】

2

【従来の技術】コンピュータシステムに使用される高性能な二次記憶システムとして、ディスクアレイシステムが知られている。
【０００３】ディスクアレイシステムは、複数の物理記憶装置をアレイ状に配置し、各物理記憶装置にデータを分割して格納しておき、前記各物理記憶装置を並列に動作させて、前記各物理記憶装置に分割して格納されるデータのリード／ライトを高速に行うことを可能としたシステムである。
【０００４】ディスクアレイシステムに関する従来技術として、例えば、D.A.Patterson, G.Gibson, and R.H.Kats, "A Case for Redundant Arrays of Inexpensive Disks（ディスクアレイ）" (in Proc. ACM SIGMOD, pp. 109-116, June 1988) 等に記載された技術が知られている。この従来技術は、冗長性を付加したディスクアレイシステムに対し、その構成に応じてレベル１からレベル５の種別を与えておくというものである。また、これらの種別に、冗長性無しのディスクアレイシステムを加え、これをレベル０と呼ぶこともある。前述の各レベルは、冗長性等に応じて異なる構成として実現するため、コストや性能特性等が異なる。そして、ディスクアレイシステムを構築するにあたって複数のレベルのアレイ（物理記憶装置の組）を混在させることも多い。ここでは、冗長性を付加したディスクアレイの組をパリティグループと呼ぶ。また、物理記憶装置についても性能や容量等によりコストが異なり、ディスクアレイシステムを構築するにあたって最適なコストパフォーマンスを実現するために、やはり性能や容量の異なる複数種の物理記憶装置を用いることがある。
【０００５】ディスクアレイシステムに格納されるデータは、前述のような前記物理記憶装置に分散して配置される。このため、ディスクアレイシステムは、ディスクアレイシステムに接続されるホストコンピュータがアクセスする論理記憶領域と前記物理記憶装置の記憶領域を示す物理記憶領域の対応付け、すなわち、アドレス変換を行う必要がある。
【０００６】アドレス変換の処理を行うディスクアレイシステムに関する従来技術として、例えば、特開平９－２７４５４４号公報等に記載された技術が知られている。この従来技術は、ホストコンピュータからの論理記憶領域に対する１／０アクセスについての情報を取得する手段と、論理記憶領域の物理記憶領域への対応付けを変更して物理的再配置を行う手段とにより、格納されたデータの最適配置を実現するというものであり、この公報には、論理記憶領域への１／０アクセス占有時間情報を、ディスクアレイ制御装置が取得する技術が開示されている。
【０００７】論理記憶領域への１／０アクセス占有時間情報は、ディスクアレイシステムの負荷分散を行うために、論理記憶領域の物理記憶領域への対応付けを変更し

て物理的再配置を行う際に、元になるデータとなるため重要である。

【０００８】

【発明が解決しようとする課題】前述した公報に記載された従来技術は、論理記憶領域への１／Ｏによる占有時間をディスクアレイ制御装置が取得するというものであるが、この従来技術に示された方法は、次に説明するような問題点を有している。

【０００９】まず、ある論理記憶領域にデータの書き込み（ライト）が行われた場合を考える。この場合、当該論理記憶領域に対応する物理記憶領域にデータのライトが行われる。物理記憶領域は、物理記憶装置内に有り、物理記憶装置は、主に物理記憶領域制御部、データをキャッシュするキャッシュメモリ及び物理記憶領域により構成される。そして、データを物理記憶領域にライトする場合、物理記憶領域制御部がライトデータをキャッシュに書込んだ時点で書き込み終了の応答がディスクアレイ制御装置に通知される。このため、前述した従来技術は、実際にデータをライトするために物理記憶領域にアクセスした時間が判らないという問題点を有することになる。

【００１０】次に、ある論理記憶領域にデータの読み込み（リード）が行われた場合を考える。この場合、当該論理記憶領域に対応する物理記憶領域にあるデータのリードが行われるが、実際には、物理記憶装置内のキャッシュメモリにそのリードデータがあった場合、物理記憶領域にはアクセスせず、キャシュメモリにアクセスし、そのデータを返す。このため、前述した従来技術は、データリードのために実際に物理記憶領域にアクセスがあったかどうかを判別することができず、また、物理記憶領域にアクセスした正確な時間が判らないという問題点を有することになる。

【００１１】また、Ａ、Ｂ、Ｃ、Ｄと複数回にわたってある論理記憶領域にアクセスがあった場合を考える。そして、Ａの応答をＡ’、Ｂの応答をＢ’等とし、アクセスＡの時刻をＡ（ｔ）、アクセスＢの時刻をＢ（ｔ）、応答Ａ’の時刻をＡ’（ｔ）等とする。ここで、アクセスＡは、データリードで物理記憶装置にアクセスしてデータをリードし、アクセスＢは、データリードでデータがキャッシュにあるため物理記憶装置にはアクセスしないでその応答Ｂ’があったものと仮定する。この場合、先にあったアクセスＡの応答Ａ’よりも後からきたアクセスＢの応答Ｂ’の方が先になる。すなわち、Ａ（ｔ）＜Ｂ（ｔ）、Ａ’（ｔ）＞Ｂ’（ｔ）となる。このとき、論理記憶領域への１／Ｏによる占有時間をディスクアレイ制御装置において取得する従来技術は、どの１／Ｏが物理記憶装置内の物理記憶領域にどのくらいアクセスしたか、あるいは、物理記憶装置のキャッシュにヒットしたか否かをしることができないという問題点を生じる。

【００１２】本発明の目的は、前述した従来技術の問題点を解決し、物理記憶装置で論理記憶領域の占有時間を取得することによって、ディスクアレイ制御装置のみでは取得不可能なシステム構成で各論理記憶領域の占有時間（実稼働時間）を取得できるようにしたストレージサブシステムを提供し、かつ、これに使用する記憶装置を提供することにある。

【００１３】また、本発明の目的は、１つの１／Ｏ毎の物理記憶装置への影響を考慮することによって、前記ディスクアレイシステムの論理記憶領域の利用率の解析や利用率予測の誤差を小さくすることができ、より最適な性能チューニングを行うために、より精度の高い物理記憶装置への１／Ｏ毎のアクセス占有時間情報を取得することができるストレージサブシステムを提供し、かつ、これに使用する記憶装置を提供することにある。

【００１４】

【課題を解決するための手段】本発明によれば前記目的は、１または複数の計算機に接続され、複数の物理記憶装置と、これらの複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とを有するストレージサブシステムにおいて、前記複数の物理記憶装置の使用状況情報を取得する手段と、前記計算機がリード／ライト対象とする論理記憶領域と前記物理記憶装置の物理記憶領域との対応付けを行う手段とが、前述複数の物理記憶装置を制御する制御装置内に設けられ、前記制御装置が、さらに、物理記憶装置の論理記憶領域と物理記憶装置の物理記憶領域との対応付けを行った情報を前記複数のそれぞれの物理記憶装置に送信する手段を備え、前記複数の物理記憶装置のそれぞれが、物理記憶領域制御装置を備え、該物理記憶領域制御装置が、物理記憶領域の使用状況を取得する手段を有することにより達成される。

【００１５】また、前記目的は、ストレージサブシステムを構成する物理記憶制御装置において、物理記憶領域制御装置を備え、該物理記憶領域制御装置が、物理記憶領域の使用状況を取得する手段、取得した物理記憶領域の使用状況情報を格納する手段、前記制御装置より受信する物理記憶領域の使用状況情報の取得要求に応じて、自物理記憶装置の物理記憶領域の使用状況情報を前記制御装置に送信する手段、及び、前記制御装置より受信する物理記憶装置の論理記憶領域と物理記憶領域とを対応付けした情報を格納する手段を備えることにより達成される。

【００１６】

【発明の実施の形態】以下、本発明によるストレージサブシステム及びそのシステムに使用する記憶装置の実施形態を図面により詳細に説明する。

【００１７】図１は本発明によるストレージサブシステムを備えた計算機システムの構成を示すブロック図、図

２は物理記憶装置の構成を示すブロック図である。図１、図２において、１０はホスト、１２はストレージサブシステム、１３はディスクアレイ制御装置、１４はディスクアレイ制御情報、１５は物理記憶装置、１６はディスクアレイ、１７は制御端末、１８はＩ／Ｏバス、１９はネットワーク、２２は物理記憶領域制御装置、２３は物理記憶領域、１３０はリード／ライト処理部、１３１は使用状況情報取得処理部、１３２は再配置判断処理部、１３３は再配置実行処理部、１４１は論理／物理対応情報、１４２はクラス構成情報、１４３はクラス属性情報、１４４は論理領域使用状況情報、１４５は物理領域使用状況情報、１４６は再配置判断対象期間情報、１４７は再配置実行時刻情報、１４８は未使用領域情報、１４９は再配置情報、１４Ａは記憶装置占有時間情報、２２１は物理記憶領域制御部、２２２はキャッシュメモリ、２２３は稼動情報取得部、２２４はデータ処理制御部、２２５はアクセス要求情報記憶部、２２６は構成・分類情報管理部、２２７は占有時間集計部である。

【００１８】図１に示す計算機システムは、上位の計算機である１または複数のホスト１０、ストレージサブシステム１２、制御端末１７から構成される。ホスト１０は、ストレージサブシステム１２にＩ／Ｏバス１８で接続され、ストレージサブシステム１２に対してデータのリードやライト処理のためのＩ／Ｏを発行する。このＩ／Ｏを行う際、ホスト１０は、ストレージサブシステム１２の論理的な記憶領域を指定する。すなわち、ホスト１０は、ストレージサブシステム内のデータに対して、通常論理的な記憶領域のアドレスによりアクセスを行う。また、Ｉ／Ｏバス１８は、例えば、ＥＳＣＯＮ、ＳＣＳＩ、ファイバチャネル等により構成される。

【００１９】ストレージサブシステム１２は、ディスクアレイ制御装置１３及び複数の物理記憶装置１５から構成される。ディスクアレイ制御装置１３は、リード／ライト処理部１３０と、使用状況情報取得処理部１３１と、再配置判断処理部１３２と、再配置実行処理部１３３とを備え、これらの処理部が、リード／ライト処理、使用状況情報取得処理、再配置判断処理、再配置実行処理等の処理を行う。また、ストレージサブシステム１２のディスクアレイ制御部は、論理記憶領域／物理記憶領域対応情報１４１、クラス構成情報１４２、クラス属性情報１４３等のディスクアレイ構成情報１４００と、論理領域使用状況情報１４４、物理領域使用状況情報１４５等の記憶装置占有時間情報１４Ａと、再配置判断対象期間情報１４６と、再配置実行時刻情報１４７と、未使用領域情報１４８と、再配置情報１４９等を保持している。なお、前述したディスクアレイ構成情報１４には、前述した情報の他、パリティグループ情報やＲＡＩＤレベル情報等が含まれてもよい。

【００２０】また、ホスト１０、ディスクアレイ制御装置１３及び制御端末１７は、相互にネットワーク１９により接続されている。ネットワーク１９は、例えば、イーサネット（登録商標）、ＦＤＤＩ、ファイバチャネル等により構成されてよい。制御端末１７は、通常、ストレージサブシステム１２の保守・管理等を行うために使用される。

【００２１】また、ホスト１０、ディスクアレイ制御装置１３及び制御端末１７には、それぞれでの処理を行うためのメモリ、ＣＰＵ等の計算機において必ず存在する構成要素をそれぞれ存在するが、本発明の実施形態の説明においては重要でないため、ここでは明記しない。

【００２２】前述のストレージサブシステム１２内に設けられる複数の物理記憶装置１５は、物理記憶装置の性能毎にクラス分けされて、クラス毎にディスクアレイ１６を構成している。また、ここでは、明示的に示していないが、複数の物理記憶装置を使用して、パリティグループが構成されている。そして、物理記憶装置１５のそれぞれは、図２に示すように、物理記憶領域２３とこの物理記憶領域２３を制御する物理記憶領域制御装置２２とにより構成され、物理記憶領域２３には、様々なデータが格納されている。

【００２３】また、前述したように、ホスト１０からは物理記憶領域２３のアドレスは直接見えてはおらず、ホスト１０は、複数の物理記憶領域２３上にある複数の論理的な記憶領域上にあるデータにアクセスを行う。すなわち、ホスト１０は、ストレージサブシステム１２内の各物理記憶装置１５の記憶領域にあるデータに対して、論理記憶領域を指定してアクセスを行う。

【００２４】ディスクアレイ制御装置１３は、複数の物理記憶装置１５と接続されており、複数の物理記憶装置１５を制御したり、前記ホスト１０から発せられたリードやライト処理命令Ｉ／Ｏを、指定のデータが存在する論理記憶領域のアドレスとその論理記憶領域のアドレスがある物理記憶領域のアドレスとを対応させて、適当な物理記憶装置１５にデータＩ／Ｏを送信し、ライト処理であれば、ホスト１０から送信されてくるデータを物理記憶装置１５に送信し、リード処理であれば物理記憶装置１５から送信されてくるデータを受信してホスト１０に送信する等の処理を行っている。

【００２５】物理記憶装置１５内に備えられる物理記憶領域制御装置２２は、物理記憶領域制御部２２１とキャッシュメモリ２２２とにより構成されている。キャッシュメモリ２２２は、物理記憶領域２３に比べデータのリード／ライトの処理の速度が速い。そして、キャッシュメモリ２２２は、ディスクアレイ制御装置１３から送信されてくるリードまたはライト命令に関するデータに関して次のように使用される。すなわち、ライト処理の場合、ディスクアレイ制御装置１３から送信されてくるライトデータが物理記憶領域２３に書き込まれる際に、データは、キャッシュメモリ２２２にも書き込まれる。また、リード処理の場合、物理記憶領域２３からデータ読

み出される際に、読み出されたデータは、キャッシュメモリ２２２に書き込まれ、あるいは、以前のリード処理によって同一のデータがキャッシュメモリに有り、そのデータに対してディスクアレイ制御装置１３からリード命令として物理記憶装置にきた場合に、そのデータを物理記憶領域２３から読み込まず、キャッシュメモリ２２２から読み込む。これにより、物理記憶装置１５の処理性能を上げることができる。

【００２６】物理記憶領域制御部２２１は、主に稼動情報取得部２２３とデータ処理制御部２２４とを備えて構成されている。データ処理制御部２２４は、ディスクアレイ制御装置１３から送信されてくるデータのリードまたはライト命令を受信する。そして、データ処理制御部２２４は、受信した命令がリード命令であった場合、キャッシュメモリ２２２にアクセスし、そのリードデータがキャッシュメモリ２２２に存在すれば、キャッシュメモリ２２２からそのリードデータを読み出し、キャッシュメモリ２２２にそのデータがなければ、物理記憶領域２３にアクセスしてそのリードデータを読み出して、ディスクアレイ制御装置１３にデータを送信する。また、データ処理制御部２２４は、受信した命令がライト処理命令であった場合、ディスクアレイ制御装置１３から送信されてくるデータをキャッシュメモリ２２２に書き込むと同時に、またはその後で、そのデータを物理記憶領域２３に書き込む。ライトデータは、キャッシュメモリ２２２に書き込まず、直接物理記憶領域２３に書き込んでもよい。

【００２７】稼動情報取得部２２３は、アクセス要求情報記憶部２２５、構成・分類情報管理部２２６、占有時間集計部２２７等により構成される。前述のデータ処理制御部２２４は、Ｉ／Ｏ処理によって指定されたデータのある論理記憶領域や、そのデータが存在する物理記憶領域２３、あるいは、キャッシュメモリ２２２にアクセスしたときに、そのアクセスの時間情報をＩ／Ｏ処理の処理種別毎（ランダムアクセスかシーケンシャルアクセスか等）に分類して占有時間集計部２２７に記録する。また、データ処理制御部２２４は、ディスクアレイ制御装置１３よりディスクアレイ内の論理記憶領域のアドレスと物理記憶領域２３のアドレスとの対応情報や物理記憶装置１５の性能等の情報を受信し、構成・分類情報管理部２２６に記録する。さらに、データ処理制御部２２４は、複数のＩ／Ｏ処理を受け付けることが可能なように、ディスクアレイ制御装置１３等から送信されてくるデータのリードまたはライト命令データ等を受信し、アクセス要求情報記憶部２２５にその命令データを記録する。

【００２８】物理記憶装置１５は、前述したような構成を有することにより、物理記憶装置１５内でＩ／Ｏ処理による論理記憶領域、物理記憶領域２３、あるいは、キャッシュメモリ２２２にアクセスしたときのアクセス時の時間情報をＩ／Ｏ処理の処理種別毎（ランダムアクセスかシーケンシャルアクセスか等）に分類して占有時間集計部２２７に記録することが可能となり、Ｉ／Ｏ処理によって、物理記憶領域にどのくらいの時間アクセスしたか、あるいは、物理記憶装置１５のキャッシュメモリ２２にヒットしたかを分類して、その占有時間の集計を行うことが可能となる。

【００２９】図３はストレージサブシステムが起動されたときのディスクアレイ制御装置の処理動作を説明するフローチャートであり、以下、これについて説明する。

【００３０】（１）ストレージサブシステム１２の始動時、ディスクアレイ制御装置１３は、自装置１３と接続されている物理記憶装置１５に対して、物理記憶装置１５内の物理記憶領域２３にある論理記憶領域のアドレスと実際にその論理記憶領域が存在する物理記憶領域のアドレスとの対応付け情報である論理／物理対応情報１４１、クラス構成情報１４２、クラス属性情報１４３等のディスクアレイ構成情報１４を送信する（ステップ３００、３１０）。

【００３１】（２）次に、ディスクアレイ制御装置１３は、前述した情報の送信により、物理記憶装置１５がアクセス可能となったときに、物理記憶領域制御装置２２から送られてくる物理記憶装置１５がアクセス可能なレディ状態に遷移した通知を受信する。ことのき、物理記憶装置１５は、ディスクアレイ構成情報１４による初期化終了の状態となっている（ステップ３２０）。

【００３２】（３）続いて、ディスクアレイ制御装置１３は、Ｉ／Ｏバス１８経由でホスト１０よりストレージサブシステム１２に、そのストレージサブシステム１２内の論理記憶領域に対してリードやライト処理のホストＩ／Ｏを送信してきたものや、ディスクアレイ制御装置同士で命令やデータを受け渡すもの等の様々なデータを受信する（ステップ３３０）。

【００３３】（４）前記受信データとしてホストＩ／Ｏを受信した場合、ディスクアレイ制御装置１３は、ホストＩ／Ｏにより指定された論理記憶領域に対するリードまたはライト要求を受信し、その論理記憶領域のアドレス（論理アドレス）を物理記憶領域のアドレス（物理アドレス）に変換する論理／物理対応情報１４１を用いて、その論理記憶領域アドレスと対応する物理記憶領域２３のアドレスを求める（ステップ３４０、３５０）。

【００３４】（５）ディスクアレイ制御装置１３は、所定のデータが存在する物理記憶領域のアドレスを指定し、リード処理の場合、前述の物理アドレスを有する物理記憶装置からリードデータを読出し、ホスト１０にリードデータを転送し、ライト処理の場合、ホスト１０から転送されたライトデータを受信し、その物理アドレスを持つ物理記憶装置にライトデータを転送する（３６０）。

【００３５】図４は論理記憶領域のアドレスと物理記憶領域のアドレスの対応が変化した場合のディスクアレイ制御装置の処理動作を説明するフローチャートであり、以下、これについて説明する。

【００３６】（１）ディスクアレイ制御装置１３は、物理記憶装置１５の増減やＲＡＩＤレベルの変化、論理記憶領域が現在ある物理記憶境域アドレスとは別の物理記憶領域のアドレスに移動する等によって、論理記憶領域のアドレスと物理記憶領域のアドレスの対応が変化したことを監視し、変化があった場合、再度、物理記憶装置１５に対して、物理記憶装置１５内の物理記憶領域２３にある論理記憶領域のアドレスと実際にその論理記憶領域が存在する物理記憶領域のアドレスとの対応付け情報である論理／物理対応情報１４１、クラス構成情報１４２、クラス属性情報１４３等のディスクアレイ構成情報１４を送信する（ステップ３１０１）。

【００３７】（２）次に、ディスクアレイ制御装置１３は、前述した情報の送信により、物理記憶装置１５がアクセス可能となったときに、物理記憶領域制御装置２２から送られてくる物理記憶装置１５がアクセス可能なレディ状態に遷移した通知を受信する。ことのき、物理記憶装置１５は、ディスクアレイ構成情報１４による更新終了の状態となっている（ステップ３２０１）。

【００３８】（３）その後の処理は、図３により説明したステップ３３００、３４００、３５００、３６００の場合と同様に実行される（ステップ３３０１、３４０１、３５０１、３６０１）。

【００３９】なお、前述したステップ３２０、３２０１において、ディスクアレイ制御装置１３は、が物理記憶領域制御装置２２から、物理記憶装置１５がアクセス可能状態に遷移したという情報を受信しなくてもよい。この場合、ある定まった時間後に物理記憶装置１５に対してリードやライト等のアクセス可能な状態になっていると仮定し、何らかの記憶領域へのアクセス指示がディスクアレイ制御装置１３にきた場合に、物理記憶装置１５に所定の処理を行うためのアクセス処理を行えばよい。また、所定のアクセス処理に対する応答がない場合、再度、所定の処理を行うためアクセス処理を行うか、あるいは、応答が帰ってくるまで待ち、一定時間中に応答がない場合、何らかの記憶領域へのアクセス指示を出したモジュールに対してその旨を伝える方式としてもよい。

【００４０】また、前述における論理／物理対応情報１４１は、論理記憶領域と物理記憶領域とを対応させる情報である。そして、論理アドレスは、ホスト１０が前記リード／ライト処理部１３０で用いる論理記憶領域を示すアドレスである。また、物理アドレスは、実際にデータが格納される物理記憶装置１５上の領域を示すアドレスであり、物理記憶装置番号及び物理記憶装置内アドレスからなる。記憶装置番号は、個々の物理記憶装置１５を示す。記憶装置内アドレスは、物理記憶装置１５内での記憶領域を示すアドレスである。

【００４１】図５はディスクアレイ制御装置１３が物理記憶装置１５の稼動情報取得部２２３内の情報を読み出す際のディスクアレイ制御装置１３の処理動作を説明するフローチャートであり、以下、これについて説明する。

【００４２】（１）ディスクアレイ制御装置１３は、ストレージサブシステム１２が起動された後、記憶装置占有時間情報１４Ａを初期化し、その後、接続されている複数の物理記憶装置１５にその物理記憶装置１５のアクセス占有時間情報の取得要求を送信する（ステップ３７１、３７２）。

【００４３】（２）次に、ディスクアレイ制御装置１３は、各物理記憶装置１５よりアクセス占有時間情報を受け取り、各物理記憶装置のアクセス占有時間情報を記憶装置占有時間情報１４Ａに格納する（ステップ３７３、３７４）。

【００４４】なお、前述したディスクアレイ制御装置１３のアクセス占有情報の取得のタイミングは、ホストＩ／Ｏやバックアップ等によるその他のモジュールから物理記憶装置２３に対するアクセスによるアクセス占有情報を各物理記憶装置１５内の占有時間集計部２２７から一定時間間隔に読み出す方式や、他のモジュール（例えばホスト１０や制御端末１７）からアクセス占有時間情報取得要求がディスクアレイ制御装置１３に送信された際等、様々であり、設計に依存する。

【００４５】前述により取得されたアクセス占有時間情報は、ディスクアレイ制御装置１３内の占有時間集計テーブルに記録される。

【００４６】図６は物理記憶装置１５内の物理記憶領域制御装置２２の処理動作を説明するフローチャートであり、次に、これについて説明する。

【００４７】（１）物理記憶領域制御装置２２は、ストレージサブシステム１２の始動時にディスクアレイ制御装置１３より送信されるデータ、すなわち、物理記憶装置１５の物理記憶領域２３にある論理アドレスと物理アドレスとの対応等の情報であるディスクアレイ構成情報１４を受信する（ステップ４００、４０１）。

【００４８】（２）ディスクアレイ構成情報１４を受信した物理記憶領域制御装置２２は、その情報を元に、稼動情報取得部２２３内の構成・分類情報管理部２２６の構成・分類情報管理テーブルや占有時間集計部２２７の占有時間集計テーブルの作成初期化を行う（ステップ４０２）。

【００４９】（３）ステップ４０２でのテーブルの初期化処理が終了すると、この物理記憶装置１５にアクセス可能であることを認識させるためディスクアレイ制御装置１３に初期化処理が終了したことを通知する。なお、ディスクアレイ制御装置１３に物理ディスク装置がアクセス可能な状態に遷移した情報を送信しなくてもよい。この場合、定まった時間後に物理記憶装置１５の物理記

憶領域制御装置２２に対して何らかの記憶領域へのアクセス指示がきた場合、物理記憶領域制御装置２２は、所定の処理を行うために物理記憶領域２３にアクセス可能な状態であれば、その物理記憶領域２３にアクセスし所定の処理を行い、不可能であれば、アクセス要求情報記憶部２２５にアクセス情報を格納し、物理記憶領域にアクセス可能になった状態で所定の処理を行うか、あるいは、所定の処理を行うために物理記憶領域２３にアクセス可能な状態でになるまで、記憶領域への何らかのアクセス指示を受け付けないようにしてもよい（ステップ４０３）。

【００５０】（４）その後、物理記憶領域制御装置２２は、ディスクアレイ制御装置１３からホストＩ／Ｏや、物理記憶装置稼動情報取得要求命令、あるいは、新たなディスクアレイ構成情報が送信されてくるのを待ってそれを受信する（ステップ４０４）。

【００５１】（５）ステップ４０４で、ディスクアレイ制御装置１３からホストＩ／Ｏを受信すると、そのＩ／Ｏがリード処理かライト処理かを判定し、リード処理であった場合、データ処理制御部２２４は、読み出すべきデータがキャッシュメモリ２２内に存在するか否かをチェックする（ステップ４０５、４０６）。

【００５２】（６）ステップ４０６のチェックで、そのデータがキャッシュメモリ２２内に存在した場合、そのデータをキャッシュメモリ２２から読み出し、また、そのデータがキャッシュメモリ２２内に存在しなかった場合、物理記憶領域２３からそのデータを読み出して、データをディスクアレイ制御装置１３に転送する（ステップ４０７、７０９、４０８）。

【００５３】（７）ステップ４０５で、ホストＩ／Ｏがライト処理であると判定された場合、データ処理制御部２２４は、ホスト１０から転送されたライトデータを受信し、キャッシュメモリ２２にそのライトデータを書き込む（ステップ４１０、４１１）。

【００５４】（８）そして、データ処理制御部２２４は、データ書き込み終了通知をディスクアレイ制御装置１３に通知すると共に、前述のライトデータを物理記憶領域２３に格納する（ステップ４１２、４１３）。

【００５５】（９）ステップ４０８の処理後、または、ステップ４１３の処理後、データ処理制御部２２４は、キャッシュメモリ２２にアクセスたか、あるいは、物理記憶領域２３にアクセスしたかの情報、ランダムリードかシーケンシャルリードか等のＪＯＢ種別情報、ライトデータを物理記憶領域２３に書き込む際のランダムリード／ライトかシーケンシャルリード／ライトか等のＪＯＢ種別情報のアクセス種別を認識し、アクセス種別毎に、キャッシュメモリ２２あるいは物理記憶領域２３にアクセスした占有時間情報を稼動情報取得部２２３内の占有時間集計部２２７に格納する（ステップ４１４、４１５）。

【００５６】（10）ステップ４０４でディスクアレイ制御装置１３から新たなディスクアレイ構成情報を受信すると、データ処理制御部２２４は、稼動情報取得部２２３の構成・分類情報管理部２２６内の情報を新たなディスクアレイ構成情報に対応して書き換える（ステップ４１８、４１９）。

【００５７】（11）ステップ４０４でディスクアレイ制御装置１３から物理記憶装置稼動情報取得要求命令を受信した場合、データ処理制御部２２４は、稼動情報取得部２２３内の占有時間集計部２２７に格納している物理記憶装置１５のアクセス占有時間情報を読み出して、それをディスクアレイ制御装置１３に送信する（ステップ４１６、４１７）。

【００５８】なお、前述したステップ４１７の処理での占有時間情報の物理記憶装置１５からディスクアレイ制御装置１３への送信は、一定時間間隔で物理記憶装置１５からディスクアレイ制御装置１３に自動的に行うようにしてもよい。この場合、ディスクアレイ制御装置１３から前記物理記憶装置稼動情報取得要求命令が物理記憶装置１３に送信されてくることはない。

【００５９】図７はディスクアレイ制御装置１３内に保持されている論理記憶領域のアドレスと物理記憶領域のアドレスとの対応を管理するためのテーブル内の論理／物理対応情報１４１の構成例を説明する図である。

【００６０】ディスクアレイ制御装置１３は、接続されている複数の物理記憶装置１５内の物理記憶領域２３にある論理記憶領域のアドレスとその論理記憶領域内の物理記憶領域のアドレスとの対応を管理している。これに使用する論理／物理対応情報１４１は、図７に示すように、特定の論理記憶領域に対して付与される論理記憶領域番号５００、論理アドレス５１０、その論理記憶領域がある物理的な記憶領域を持つ記憶装置番号５２１と物理的な記憶領域のアドレス５２２とによる物理アドレス５２０、その物理記憶装置１５の性能を示すレイドレベル５３０、その物理記憶装置１５が属しているパリティグループ番号５４０のそれぞれが対応付けられて構成される。ディスクアレイ制御装置１３は、このような論理／物理対応情報１４１を有することにより、ホスト１０や制御端末１７、その他のモジュール（例えば、その他のディスクアレイ制御装置等）からリードやライト等の処理が論理記憶領域のアドレスを指定して自ディスクアレイ制御装置１３に対してアクセスされた場合、論理領域のアドレスを物理的な記憶領域のアドレスに変換して、データのリード／ライト処理を物理記憶装置１５に対して正確に行うことができる。

【００６１】図８はディスクアレイ制御装置１３内に格納される論理領域使用状況情報１４４と物理領域使用状況情報１４５等の記憶装置占有時間情報１４１の例を示す図である。これらの情報は、占有時間集計テーブルとして構成されている。

【００６２】ディスクアレイ制御装置１３は、ホストＩ／Ｏやバックアップ等によるその他のモジュールからのアクセスによる物理記憶装置１５の物理記憶領域２３へのアクセス占有情報を各物理記憶装置１５内の占有時間集計部２２７から定期的に読み出して、そのアクセス占有時間情報を受信ディスクアレイ制御装置１３内の占有時間集計テーブルに記録する。図８に示す例は、論理記憶領域番号６０１毎、Ｉ／Ｏ　ＪＯＢ種別６０２毎に占有時間を集計したものである。Ｉ／Ｏ　ＪＯＢ種別として、図示例では、シーケンシャルリード６１０、シーケンシャルライトデータ６２０、シーケンシャルライトパリティ６３０、ランダムリード６４０、ランダムライトパリティ６６０、キャッシュヒット時６７０、合計６８０が示されているが、さらに他のＩ／Ｏ　ＪＯＢ種別があってもよい。

【００６３】物理記憶装置１５から読み出される占有時間は、前述に限らず、Ｉ／Ｏ毎のアクセス占有時間の累積値や、ユニバーサルな時間値、マシン固有な時間値による記録であってもよい。また、ディスクアレイ制御装置１３において、各論理記憶領域や物理記憶領域のアクセス占有時間を編集して、物理記憶装置毎やパリティグループ毎のアクセス占有時間を求めた値に基づいて新たに占有時間情報テーブルを作成してもよい。

【００６４】前述により、ホスト１０や制御端末１７等が物理記憶装置の稼動情報取得要求をディスクアレイサブシステム１２に出した場合、ホスト１０や制御端末１７等は、直接物理記憶装置１５にアクセスして論理記憶領域や物理記憶領域のアクセス占有時間情報を取得しなくてもディスクアレイ制御装置１３からその論理記憶領域や物理記憶領域のアクセス占有時間情報を取得することが可能となる。

【００６５】図９は物理記憶装置１５内の稼動情報取得部２２３内の構成・分類情報管理部２２６に格納される論理記憶領域のアドレスと物理記憶領域のアドレスとの対応付けを管理するテーブルの構成例を示す図である。

【００６６】物理記憶装置１５内のデータ処理制御部２２４は、ディスクアレイ制御装置１３より、ストレージサブシステム１２の始動時や、物理記憶装置の増減やレイドレベルの変化、論理記憶領域の移動等によって生じる論理記憶領域のアドレスと物理記憶領域のアドレスとの関係が変化した際に、その論理記憶領域と物理記憶領域との対応付け情報を受信し、それを構成・分類情報管理部２２６に図９に示すような対応テーブルに格納する。この対応テーブルは、特定の論理記憶領域に対して付与される論理記憶領域番号７００、論理アドレス７１０、その論理記憶領域がある物理的な記憶領域を持つ記憶装置番号７２１と物理的な記憶領域のアドレス７２２とによる物理アドレ７５２０により構成される。これにより、物理記憶装置１５は、どのアドレス論理記憶領域が自物理記憶装置１５内の物理記憶領域アドレスのどこ

にあるかを認識することができる。

【００６７】図１０は物理記憶装置１５内の稼動情報取得部２２３の占有時間集計部２２７に累積・格納されるＩ／Ｏによる記憶領域へのアクセスによる占有時間情報のテーブルの構成例を示す図である。

【００６８】このテーブルは、論理記憶領域番号８０１毎、Ｉ／Ｏ　ＪＯＢ種別８０２毎に占有時間を集計したものである。Ｉ／Ｏ　ＪＯＢ種別として、図示例では、シーケンシャルリード８１０、シーケンシャルライトデータ８２０、シーケンシャルライトパリティ８３０、ランダムリード８４０、ランダムライトパリティ８６０、キャッシュヒット時８７０、合計８８０が示されているが、さらに他のＩ／ＯＪＯＢ種別があってもよい。

【００６９】物理記憶装置１５内のデータ処理制御部２２４は、ホストからのＩ／Ｏ等により記憶装置にアクセスがあると、アクセスがあった各論理記憶領域８０１に関して、アクセスのＪＯＢ種別８０２毎に、占有時間８９０を占有時間集計部２２７に累積する。これにより、物理記憶装置１５内の論理記憶領域の番号と、その論理記憶領域へのアクセス種別による占有時間と、ある時間内の合計占有時間との関係を物理記憶装置１５内で得ることが可能となる。

【００７０】前述した本発明の実施形態によれば、Ｉ／Ｏ毎の記憶装置の占有時間情報を取得することができ、かつ、記憶領域へのアクセスによる記憶装置の占有時間情報の取得を物理記憶装置内で実現することができる。

【００７１】本発明の実施形態によれば、前述により、ストレージサブシステムを構成する複数の物理記憶装置のそれぞれが、物理記憶装置の論理記憶領域の占有時間を取得することが可能となり、ディスクアレイ制御装置のみでは取得不可能なシステム構成で各論理記憶領域の占有時間（実稼働時間）取得することが可能となる。また、前述した本発明の実施形態によれば、１つのＩ／Ｏ毎の物理記憶装置への影響を考慮することによって、前記ストレージサブシステムの論理記憶領域の利用率の解析や利用率予測の誤差を小さくすることができ、より最適な性能チューニングを行うために、より精度の高い物理記憶装置への１／Ｏ毎のアクセス占有時間情報の取得を行うことができる。

【００７２】

【発明の効果】以上説明したように本発明によれば、物理記憶装置が論理記憶領域の占有時間を取得することができ、ディスクアレイ制御装置のみでは取得不可能なシステム構成で各論理記憶領域の占有時間（実稼働時間）を取得することができる。

【００７３】また、本発明によれば、物理記憶装置への１／Ｏ毎のアクセス占有時間情報を取得することができるため、１つのＩ／Ｏ毎の物理記憶装置への影響を考慮したストレージサブシステムの論理記憶領域の利用率の解析や利用率の予測を小さい誤差で行うことが可能とな

り、より最適なストレージサブシステムの性能のチューニングを行うことができる。

【図面の簡単な説明】

【図1】本発明によるストレージサブシステムを備えた計算機システムの構成を示すブロック図である。

【図2】物理記憶装置の構成を示すブロック図である。

【図3】ストレージサブシステムが起動されたときのディスクアレイ制御装置の処理動作を説明するフローチャートである。

【図4】論理記憶領域のアドレスと物理記憶領域のアドレスの対応が変化した場合のディスクアレイ制御装置の処理動作を説明するフローチャートである。

【図5】ディスクアレイ制御装置が物理記憶装置の稼動情報取得部内の情報を読み出す際のディスクアレイ制御装置の処理動作を説明するフローチャートである。

【図6】物理記憶装置内の物理記憶領域制御装置の処理動作を説明するフローチャートである。

【図7】ディスクアレイ制御装置内に保持されている論理記憶領域と物理記憶領域とのアドレスの対応を管理する論理／物理対応情報の構成例を説明する図である。

【図8】ディスクアレイ制御装置内に格納される論理領域使用状況情報と物理領域使用状況情報等の記憶装置占有時間情報の例を示す図である。

【図9】物理記憶装置内の構成・分類情報管理部に格納される論理記憶領域と物理記憶領域とのアドレスの対応付けを管理するテーブルの構成例を示す図である。

【図10】物理記憶装置内の稼動情報取得部の占有時間集計部に累積・格納される記憶領域へのアクセスによる占有時間情報のテーブルの構成例を示す図である。

【符号の説明】

10 ホスト

12 ストレージサブシステム
13 ディスクアレイ制御装置
14 ディスクアレイ制御情報
15 物理記憶装置
16 ディスクアレイ
17 制御端末
18 I／Oパス
19 ネットワーク
22 物理記憶領域制御装置
23 物理記憶領域
130 リード／ライト処理部
131 使用状況情報取得処理部
132 再配置判断処理部
133 再配置実行処理部
141 論理／物理対応情報
142 クラス構成情報
143 クラス属性情報
144 論理領域使用状況情報
145 物理領域使用状況情報
146 再配置判断対象期間情報
147 再配置実行時刻情報
148 未使用領域情報
149 再配置情報
14A 記憶装置占有時間情報
221 物理記憶領域制御部
222 キャッシュメモリ
223 稼動情報取得部
224 データ処理制御部
225 アクセス要求情報記憶部
226 構成・分類情報管理部
227 占有時間集計部

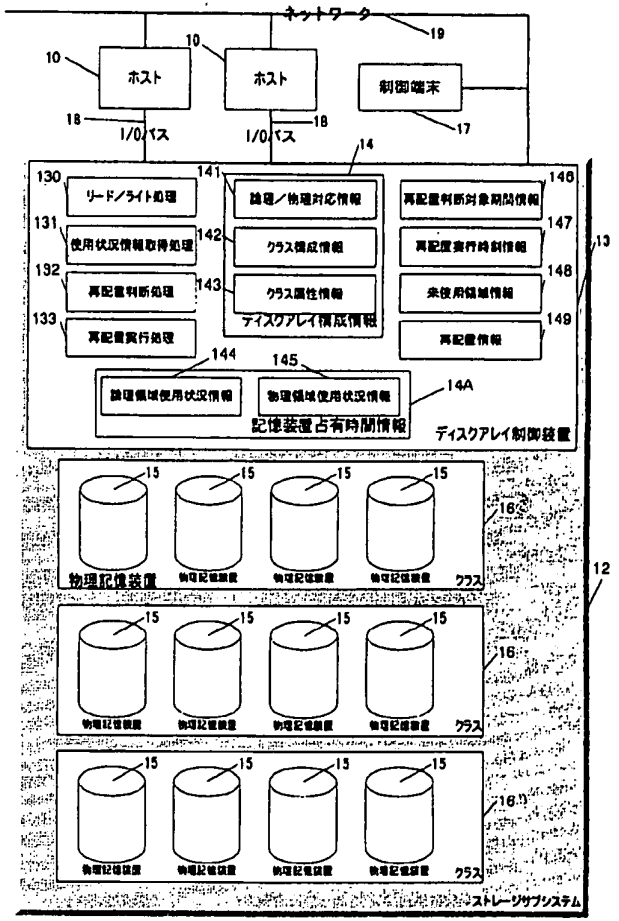
【図8】

図8

601 論理記憶領域番号 ＼ 602 I/O JOB種別	610 Sequential Read	620 Sequential Write data	630 Sequential Write Parity	640 Random Read	650 Random write Data	660 Random write Parity	670 キャッシュヒット時	680 合計
1	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
2	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
3	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
4	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
…	…	…	…	…	…	…	…	…

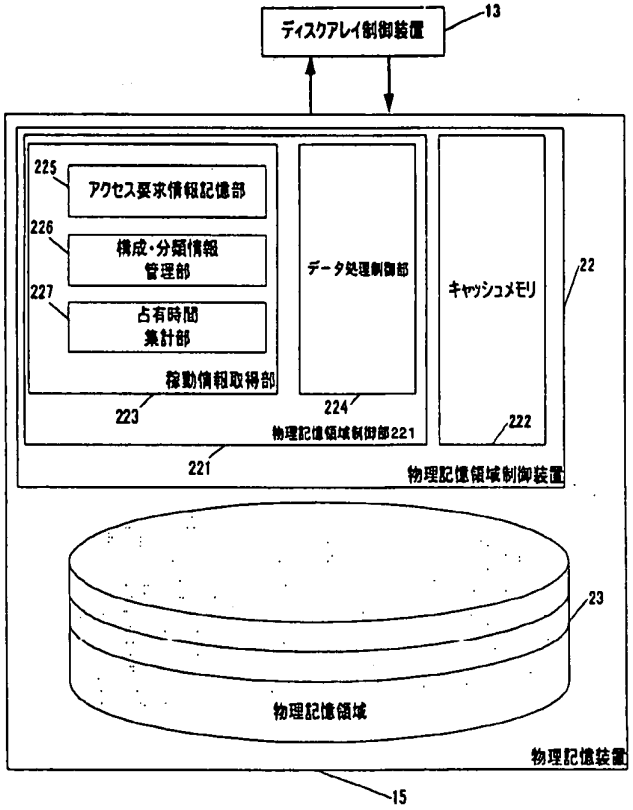
【図1】

図1



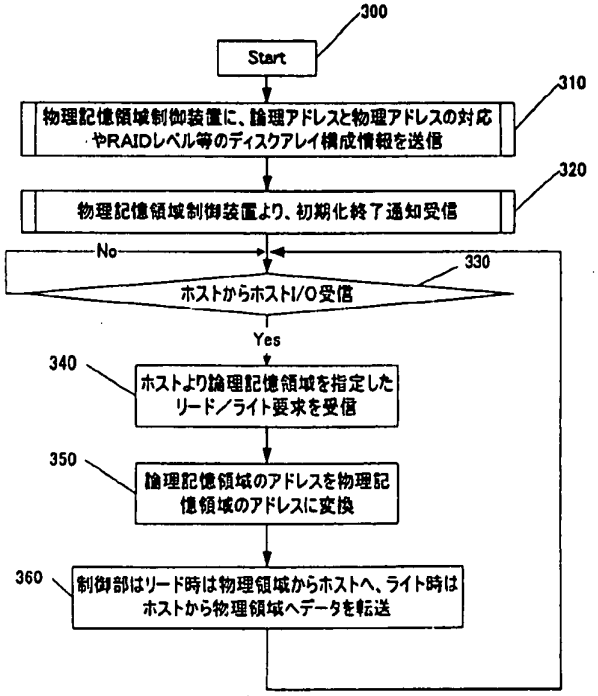
【図2】

図2



【図3】

図3



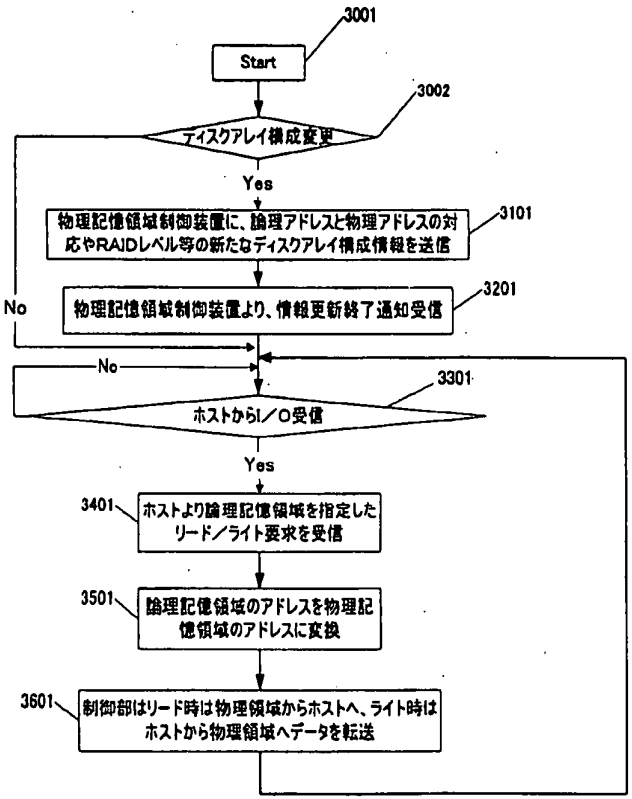
【図9】

図9

論理記憶領域番号 (700)	論理アドレス (710)	物理アドレス (720)	
		物理記憶装置番号 (721)	物理記憶領域アドレス (722)
3	3000~3999	1	0~999
4	4000~4999	1	1000~1999
5	5000~5999	1	2000~2999

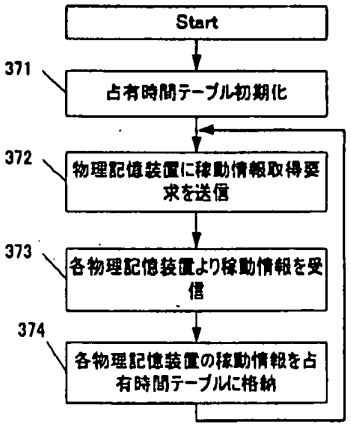
【図4】

図4



【図5】

図5



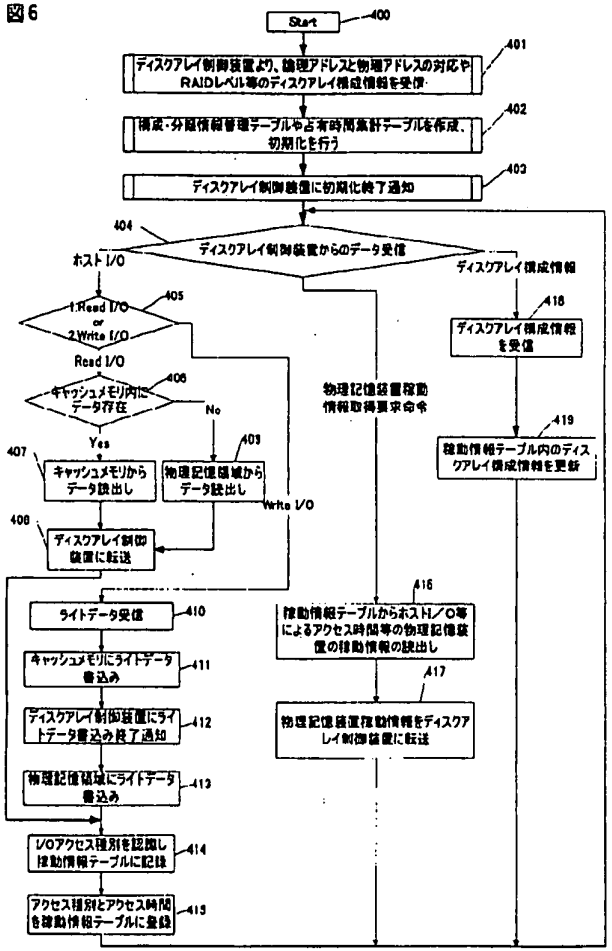
【図7】

図7

500 論理記憶領域番号	510 論理アドレス	520 物理アドレス		530 レイドレベル	540 パリティグループ番号
		521 物理記憶装置番号	522 物理記憶領域アドレス		
0	0~999	0	0~999	1	100
1	1000~1999	0	1000~1999	1	100
2	2000~2999	0	2000~2999	1	100
3	3000~3999	1	0~999	5	120
4	4000~4999	1	1000~1999	5	120
5	5000~5999	1	2000~2999	5	120
…	…	…	…	…	…

【図6】

図6



【図10】

図10

801 / 802 I/O JOB種別 / 論理記憶領域番号	810 Sequential Read	820 Sequential Write data	830 Sequential Write Parity	840 Random Read	850 Random write Data	860 Random write Parity	870 キャッシュヒット時	880 合計
3	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
4	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間
5	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間	占有時間

890

フロントページの続き

(72)発明者 荒川 敏史
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
(72)発明者 大枝 高
神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
(72)発明者 荒井 弘治
神奈川県小田原市国府津2880番地 株式会社日立製作所ストレージシステム事業部内

Fターム(参考) 5B018 GA07 MA14 QA16
5B065 BA01 BA06 CA13 CA30 CC03
CH19 ZA02

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.